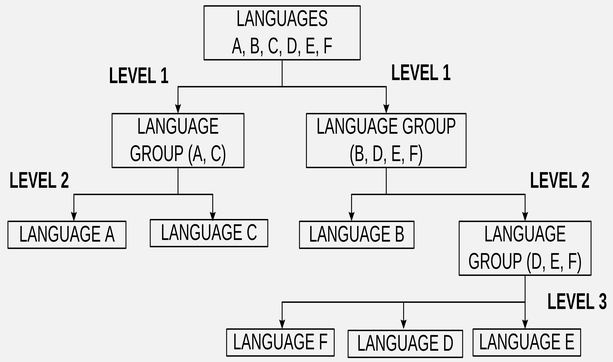
A HIERARCHICAL FRAMEWORK FOR LANGUAGE IDENTIFICATION

Saad Irtza[1,2], Vidhyasaharan Sethu[1], Haris Bavattichalil[1], Eliathamby Ambikairajah[1,2], Haizhou Li[3]
[1] The School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney NSW 2052, Australia
[2] ATP Laboratory, National ICT Australia (NICTA), Australia, Sydney NSW 2015, Australia
[3] Institute for Infocomm Research, A*STAR, Singapore

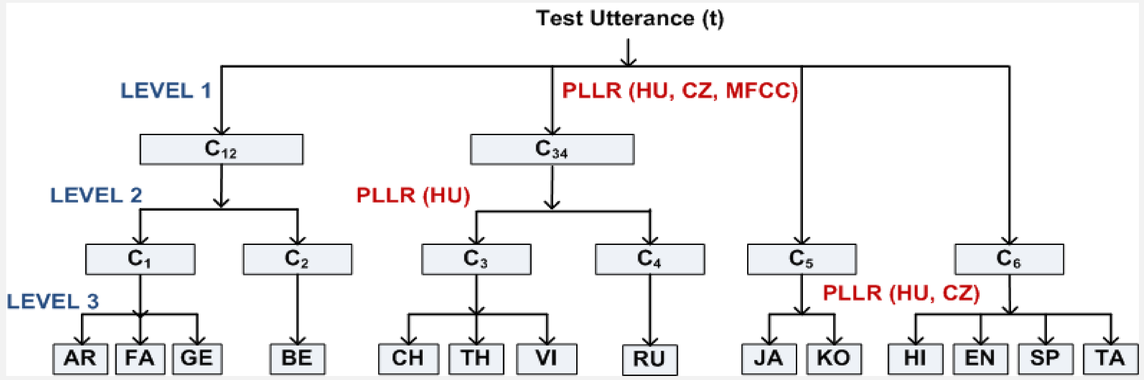
s.irtza@unsw.edu.au

Introduction

- State of art LID system:
  - Flat structure and treat all languages equally
  - Incapable of exploiting similarities between languages
- Hierarchical LID framework:
  - Based on assumption that similarities exist between languages
  - Involves series of classification at multiple levels
  - Allows to choose most effective features in each level
  - Allows target languages to be identified in final layer
- Each level of tree acts as individual LID system
- Performance comparison of flat and hierarchical LID



Hierarchical Tree Structure (NIST 2007 Database)



Hierarchical Language Clustering

- Cosine similarity score (CSS) is used as the similarity score between two individual languages (A and B):

$$S^{\phi}(A,B) = \frac{L_A^{\phi} \cdot L_B^{\phi}}{\left\|L_A^{\phi}\right\| \left\|L_B^{\phi}\right\|}$$

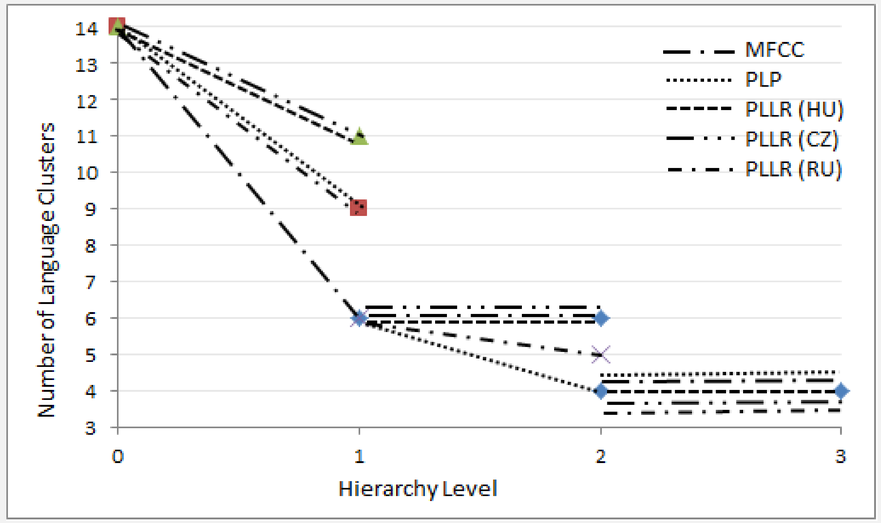
  - L_A^{ϕ}, L_B^{ϕ} i-vectors from language A and B extracted from front end ϕ
- The Unweighted Pair-Group method of Average (UPGMA) is used as similarity measure between language groups:

$$S^{\phi}(C_1,C_2) = \frac{\sum_{m \in C_1, n \in C_2} S^{\phi}(m,n)}{n_{C_1} n_{C_2}}$$

  - m, n denote languages utterances belonging to clusters C_1 and C_2 respectively;
  - n_{C_1}, n_{C_2}: total number of utterances in C_1 and C_2 respectively.
- Agglomerative hierarchical clustering algorithm is used
- Language cluster C is expanded to include language *C* iff

$$\forall_{i,j \in C_m} S^{\phi}(i,j) - \forall_{k \in C_m} S^{\phi}(k,C) < \beta$$

Selection of Features



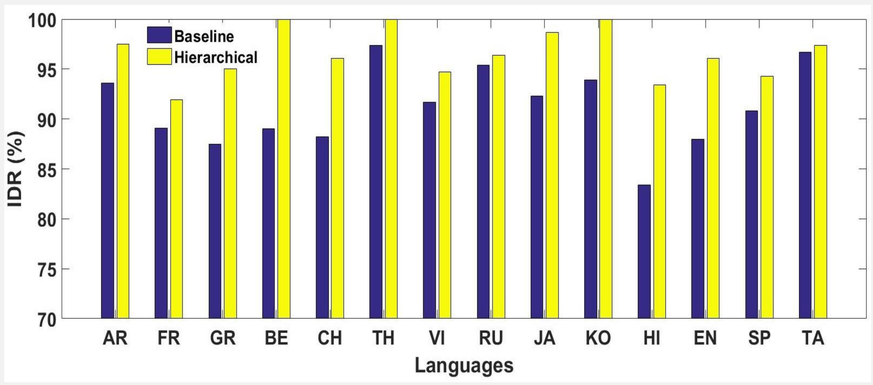
Experimental Setup

- 13 dimensional MFCCs and PLP coefficients, augmented with SDCs based on 13-7-1-3 configuration
- PLLR features of 59 (HU), 50 (RU) and 43 (CZ) dimensions augmented with SDCs based on X-1-5-1 configuration. X is original feature dimension
- Universal Background Model: 1024 component GMM
- i-vectors: 400 dimensions
- LDA is used prior to GPLDA
- LLR's computed at each level are propagated down to next hierarchy level and added to LLR's in that level
- Baseline System: Fusion of PLLR (HU, RU and CZ) front ends [1]

[1] L. F. D'Haro, et. Al., ISCA, Minneapolis, MN, USA, 2014

Results

- Performance of flat and hierarchical systems are compared in terms of identification rate and error reduction



Effect of Level wise feature selection

Level	Confusion between Clusters	Misclassifications error rate (%)		Error reduction (%)
		Baseline	HLID	
1	C_{12},C_{34},C_5,C_6	5.9	**1.9**	**4**
2	C_1,C_2	1.5	**0.62**	**0.9**
2	C_3,C_4	1.0	**0**	**1**
3	AR, GR, FR	4.2	**0.83**	**3.4**
3	CH, TH, VI	6.2	**1.5**	**4.7**
3	JA, KO	0	**0**	**0**
3	HI, EN, SP, TA	8.2	**1.4**	**6.8**
	Overall	**9.64**	**3.98**	**5.66**

Conclusion and Future Work

- A novel hierarchical framework is proposed for language identification
- The proposed hierarchical structure
  - Uses bottom up approach to find the language clusters
  - Selects a suitable front end for clustering at each level
  - Selects most discriminative features for classification at each level
- Level wise feature selection reduces misclassification at each level
- Future Work:
  - Feature selection method for classification
  - Evaluation of this framework on most recent and challenging databases